3. A gray-level digital image $\boldsymbol{X}$ of size 100×100 is cropped into small images of size 3×3 by sliding the center of a 3×3 window to every pixel of $\boldsymbol{X}$. If the center of the window is on the boundary of the image $\boldsymbol{X}$, zero padding is used to obtain images of size 3×3 so that we get 10000 small images of size 3×3. Each 3×3 image is flattened into a column vector and is expressed as $\mathbf{z}^k = (z_1^k, \dots, z_9^k, 1)^T$, $k = 1, \dots, 10000$, and is input to a typical fully-connected layer of a multilayer perceptron (MLP) with a linear activation function to generate the output vectors $\mathbf{y}^k = (y_1^k, \dots, y_6^k)^T$. The network parameters of this layer are denoted by w_{ij} and b_j, $0 < i < 10$, $0 < j < 7$.

(a) Express the outputs y_j^k in terms of the inputs z_i^k.

用输入z表示输出y。

(5 Marks)

(b) Construct the matrix $\boldsymbol{W}$ that contains all network parameters, and express the output vector $\mathbf{y}^k$ in terms of the input vector $\mathbf{z}^k$.

(5 Marks)

构造包含所有网络参数的矩阵W，用输入向量zk表示输出向量yk。

(c) 六张输出图像 Y_j，$j = 1, \dots, 6$，大小与输入图像 X 相同，由 10000 个输出向量 y^k，$k = 1, \dots, 10000$ 构成。用输入图像 X 表示输出图像 Y_j。

(c) Six output images, $\boldsymbol{Y}_j$, $j = 1, \dots, 6$, of the same size with the input image $\boldsymbol{X}$, are constructed by the 10000 output vectors $\mathbf{y}^k$, $k = 1, \dots, 10000$. Express the output images $\boldsymbol{Y}_j$ in terms of the input image $\boldsymbol{X}$.

(10 Marks)

(d) 假设该网络由100张大小为100×100的图像训练。
用于训练网络参数W 或者 wij & bi的训练样本数量是多少？

(d) Suppose this network is trained by 100 images of size 100×100. What is the number of training samples used to train the network parameters $\boldsymbol{W}$ or w_{ij} and b_j?

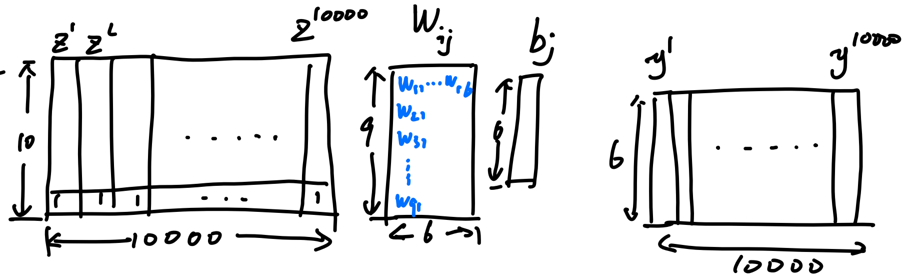
23-S2-Q2

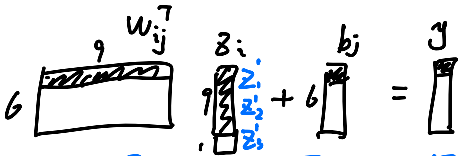
Q: $3\times3 \to 100\times100 \to 10000$ $z^k = (z_1^k\ z_2^k \cdots z_9^k, 1)^T$, $k = 1 \sim 10000$

$y^k = (y_1^k, \cdots y_6^k)^T$ w_{ij} and b_j

(a) $z_i^k \to y_j^k$?

Solution





T 标量,不能转置,也不用转置,不是矩阵

$$y_j^k = \sum_{i=1}^{9} w_{ij} z_i^k + b_j$$ $(j = 1, 2, \cdots 6)$

$$y_1^k = \sum_{i=1}^{9} w_{i1} z_i + b_1$$

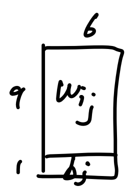
$$= w_{11} z_1 + w_{21} z_2 + \cdots w_{91} z_9 + b_1$$

(b) W? $z^k \to y^k$

Solution

we can construct the weight matrix W of size 10×6 wherer each column corresponds the the weight w_{ij} and the last one is bias b_j

$$W = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{16} \\ w_{21} & w_{22} & \cdots & w_{26} \\ \vdots & \vdots & \ddots & \vdots \\ w_{91} & w_{92} & \cdots & w_{96} \\ b_1 & b_2 & \cdots & b_6 \end{bmatrix}$$



$$y^k = W^T z^k$$

(c) $W_j = \begin{bmatrix} w_{1j} & w_{2j} & w_{3j} \\ w_{4j} & w_{5j} & w_{6j} \\ w_{7j} & w_{8j} & w_{9j} \end{bmatrix}$

$X \longrightarrow Y$ $\qquad Y_j$

1张 $\qquad$ 1张 → 6张，60000个

100×100 $\qquad$ 10000个

↓卷积：每个pixel $\qquad$ 3×3 $\quad$ W卷后，得 $3 \times 3 \to y^k$.

10000张 $\qquad$ 卷10000张，得10000个 y^k

3×3 $\qquad$ $j = 1,2,3,4,5,6$

$$\begin{array}{c|ccc} & 1 & 2 & 3 \\ \hline 1 & z_1 & z_2 & z_3 \\ 2 & z_4 & z_5 & z_6 \\ 3 & z_7 & z_8 & z_9 \end{array} \xrightarrow{v} \quad \left(\begin{bmatrix} w_1 & w_2 & w_3 \\ w_4 & w_5 & w_6 \\ w_7 & w_8 & w_9 \end{bmatrix} \boxed{b} \right) \times 6 \text{个} \to 6\text{个特征图}$$

u $\qquad$ t $\qquad$ s $\qquad$ bias

X

ref to LSI 2.2.5 example P52

$$Y_j(u,v) = \sum_{t=-1}^{1} \sum_{s=-1}^{1} W_j(s,t)\, X(u-s, v-t) + b_j$$

Y在 $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$ 上每个都卷一次

得到 9个结果

$$Y_j = X * W_j + b_j$$

(d) Each of the 100 training image of size 100×100 provide 10000 Samples

Therefore,

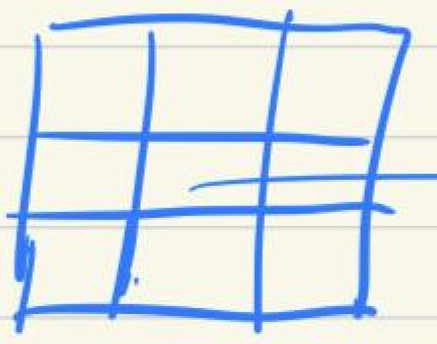
Number of training sample = 100 images $\times$ 10000 sample/image

= 1,000,000

answer 2

(c)

$$y^k = w^T \cdot z^k$$



$x(u, v)$

$$y_j^k = w_j^T z^k$$

↳ 10×1 matrix and $0 \le k \le 10000$

$$z^k(u,v) = [x(u-1,v-1), x(u,v-1), \ldots, x(u+1,v+1)]^T$$

$$Y_j(u,v) = w_j^T \cdot z^k(u,v)$$

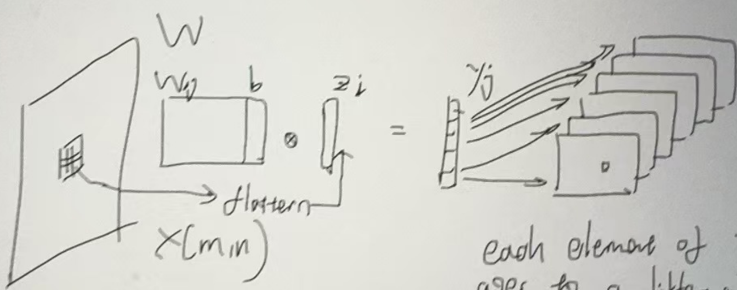
$$= w_j^T \cdot [x(u-1,v-1), \ldots, x(u+1,v+1)]^T$$

(d) each 100×100 image gives 10,000 samples.

So there are 1000,000 samples.

answer 3

$$Y_j(m,n) = \sum_{i=1}^{9} W_{ij} \cdot \text{flatten}\begin{pmatrix} x_{m-1,n-1} & x_{m-1,n} & x_{m-1,n+1} \\ x_{m,n-1} & x_{m,n} & x_{m,n+1} \\ x_{m+1,n-1} & x_{m+1,n} & x_{m+1,n+1} \end{pmatrix} + b_j$$



each element of Y_j goes to a different picture with same (m,n) coordinate.

$$W = \begin{bmatrix} W_{ij} \\ b \end{bmatrix}$$

which contains $10 \times 6 = 60$ parameters.

answer 4

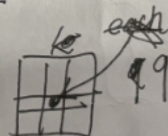
$100 \times 100 \rightarrow 10000$

So: we must make convolution on each pixel to make the pixel the same.

$Y_j, j = 1, \ldots, 6.$ ~~one Y_j, has one $1, \ldots, 10000$~~

y_i is from 9 x pixels. 10000 output y^k, ~~$1, \ldots, 10000$~~

so we can see: $Y = \sum_{-1}^{1} \sum_{-1}^{1}$ convolution kernel.



9 position + bias to get one y_i in Y.